

STABLE COMMUTATOR LENGTH IS RATIONAL IN FREE GROUPS

DANNY CALEGARI

Dedicated to Shigenori Matsumoto on the occasion of his 60th birthday.

ABSTRACT. For any group, there is a natural (pseudo-)norm on the vector space B_1^H of real homogenized (group) 1-boundaries, called the *stable commutator length* norm. This norm is closely related to, and can be thought of as a relative version of, the Gromov (pseudo)-norm on (ordinary) homology. We show that for a free group, the unit ball of this pseudo-norm is a rational polyhedron.

It follows that stable commutator length in free groups takes on only rational values. Moreover every element of the commutator subgroup of a free group rationally bounds an injective map of a surface group.

The proof of these facts yields an algorithm to compute stable commutator length in free groups. Using this algorithm, we answer a well-known question of Bavard in the negative, constructing explicit examples of elements in free groups whose stable commutator length is not a half-integer.

1. INTRODUCTION

Stable commutator length is a numerical invariant of elements in the commutator subgroup of a group. It is intimately related to (two-dimensional) bounded cohomology, and appears in many areas of low-dimensional topology and dynamics, from the Milnor-Wood inequality, to the 11/8 conjecture. However, although a great deal of work has gone into estimating or bounding stable commutator length in many contexts, there are very few known nontrivial examples of finitely presented groups in which it can be calculated exactly, and virtually no cases where the range of stable commutator length on a given group can be understood arithmetically.

The most significant results of this paper are as follows:

- (1) We show that stable commutator length in free groups takes on only rational values, and give an explicit algorithm to compute the value on any given element. This is the first example of a group with infinite dimensional second bounded cohomology group H_b^2 in which stable commutator length can be calculated exactly.
- (2) We show how to extend stable commutator length to a (pseudo)-norm on the vector space B_1^H of homogenized real (group) 1-cochains that are boundaries of 2-cochains. In the case of a free group, this is a genuine norm. We show that the intersection of the unit ball in this norm with any finite dimensional rational subspace of B_1^H of a free group is a finite-sided rational polyhedron. This invites comparison with the Thurston norm on

the 2-dimensional homology of a 3-manifold [25], although the relationship between the two cases is subtle and deserves further investigation.

- (3) We give examples of explicit elements in the commutator subgroup of F_2 (the free group of rank 2) for which the stable commutator length is not in $\frac{1}{2}\mathbb{Z}$. This answers in the negative a well-known question of Bavard [1].

We now elaborate on these points in turn.

Let G be a group. For $g \in [G, G]$, the *commutator length* of g , denoted $\text{cl}(g)$, is the smallest number of commutators in G whose product is equal to g . The *stable commutator length* of g , denoted $\text{scl}(g)$, is the limit of $\text{cl}(g^n)/n$ as $n \rightarrow \infty$. In geometric language, (see e.g. Gromov [15]) cl is sometimes called “filling genus”, and scl is called “stable filling genus”. This quantity is intimately related, by Bavard duality and an exact sequence (see Theorem 2.12 and Proposition 2.11) to the second bounded cohomology H_b^2 of G , with its Banach norm. Despite a considerable amount of research, there are very few known examples of groups G in which scl can be calculated exactly (except when it vanishes identically). This is partly because the groups H_b^2 , when nontrivial, tend to be very large in general: when G is word-hyperbolic, H_b^2 is not merely infinite dimensional, but is not even separable as a Banach space. Calculating scl is tantamount to solving an extremal problem in H_b^2 . (Technically, one solves the extremal problem in the space of *homogeneous quasimorphisms* Q , see Definition 2.10. The spaces Q and H_b^2 are both Banach spaces, and are related by the coboundary δ , which is Fredholm when G is finitely presented.)

Gromov ([15] 6.C₂) asked whether scl is always rational in finitely presented groups. The answer to Gromov’s question is known to be *no*: Dongping Zhuang gave the first examples in [26]. These examples occur in generalized Stein-Thompson groups of PL homeomorphisms of the circle, where one can show that H_b^2 is actually finite dimensional, and everything can be calculated explicitly. In this paper we show that scl is *rational* in free groups (and some closely related groups), and moreover we give an explicit algorithm to compute the value of scl on any element.

If g_1, g_2, \dots, g_m are elements in G , define $\text{cl}(g_1 + \dots + g_m)$ to be the smallest number of commutators in G whose product is equal to the product of conjugates of the g_i . Let $\text{scl}(g_1 + \dots + g_m)$ denote the limit of $\text{cl}(g_1^n + \dots + g_m^n)/n$ as $n \rightarrow \infty$. This function can be extended by linearity and continuity in a unique way to a pseudo-norm on B_1 , the vector space of real group 1-chains on G that are in the image of the boundary map $\partial : C_2 \rightarrow C_1$. This function vanishes identically on the subspace H of B_1 spanned by terms of the form $g^n - ng$ and $g - hgh^{-1}$ for $g, h \in G$ and $n \in \mathbb{Z}$, and descends to a pseudo-norm on the quotient B_1/H , or B_1^H for short. When G is hyperbolic, scl is a genuine norm on B_1^H . We show that in a free group, this scl norm is *piecewise rational linear* (denoted PQL) on finite dimensional rational subspaces of B_1^H . So for any finite set of elements $g_1, g_2, \dots, g_m \in G$, there is a uniform upper bound on the denominators of the values of scl on integral linear chains $\sum_i n_i g_i$ in B_1^H .

One should compare the scl norm with the Gromov-Thurston norm [25], which is a norm on H_2 of an irreducible, atoroidal 3-manifold, and whose most significant feature is that it is a piecewise rational linear function. In Thurston’s definition (in which one restricts to embedded surfaces) this is straightforward to show. In Gromov’s definition (in terms of chains, or immersed surfaces) this is a very deep

theorem, whose proof depends on the full power of Gabai's theory of sutured hierarchies [12], and taut foliations. In fact, it is reasonable to think of the scl norm as a relative Gromov-Thurston norm, with Gromov's definition. Our proof of rationality is conceptually close in some ways to an argument due to Oertel [21] insofar as both proofs reduce the problem of calculating the norm to a linear programming problem in the vector space of weights carried by a finite constructible branched surface. However, there are crucial differences between the two cases. In Oertel's case, the branched surface might have complicated branch locus, but it comes with an embedding in a 3-manifold. In our case, the branch locus is simple, but the branched surface is merely *immersed* in a 3-manifold. It is intriguing to try to find a natural generalization of both theories.

In his seminal paper [1] on stable commutator length, Bavard asked whether stable commutator length in free groups takes values in $\frac{1}{2}\mathbb{Z}$. There were several pieces of direct and indirect evidence for this conjecture. Firstly, where certain (geometric or homological) methods for estimating stable commutator length in free groups give exact answers, these answers in every case confirm Bavard's guess. Secondly, in the (analogous) context of 3-manifold topology, one knows that the Gromov norm of an integral 2-dimensional homology class is in $2\mathbb{Z}$ (the factor of 4 arises because Gromov norm counts triangles, whereas stable commutator norm counts handles). It was generally felt that Bavard's conjecture was eminently plausible, and it is therefore surprising that our algorithm produces many elements whose stable commutator length is not in $\frac{1}{2}\mathbb{Z}$. In fact experiments suggest that arbitrarily large denominators occur, with arbitrary prime factors. In view of these examples, the fact that stable commutator length is rational in free groups is seen to be a more delicate and subtle fact than one might have imagined, and stable commutator length to be a richer invariant than previous work has suggested.

The organization of this paper is as follows. In § 2 we state definitions and sketch proofs of background results which pertain to stable commutator length in groups in general. In § 3 we specialize to the case of free groups. The purpose of this section is to state and prove the "Rationality Theorem", whose precise statement is the following:

Rationality Theorem. *Let F be a free group.*

- (1) $\text{scl}(g) \in \mathbb{Q}$ for all $g \in [F, F]$.
- (2) Every $g \in [F, F]$ rationally bounds an extremal surface (in fact, every rational chain C in B_1^H rationally bounds an extremal surface)
- (3) The function scl is piecewise rational linear on B_1^H .
- (4) There is an algorithm to calculate scl on any finite dimensional rational subspace of B_1^H .

Similar rationality results hold for stable commutator length in virtually free groups, and fundamental groups of noncompact Seifert-fibered 3-manifolds. Finally, in § 4 we explicitly describe an algorithm for computing the stable commutator length in free groups, and discuss a simple example that answers Bavard's question in the negative.

2. BACKGROUND

For the convenience of the reader, we collect here some basic definitions and properties that will be used in subsequent sections. As general background, see [1],

[3], [9] and [14]. Note that the reference [9], although the most detailed, complete and relevant to the material in this paper, is an unfinished manuscript (which is readily available online) and therefore we have tried to refer to this manuscript by section number (which one can expect to be reasonably stable) rather than by page number.

2.1. Stable commutator length. In this section we give the definitions and basic properties of stable commutator length in groups. This is a numerical invariant of elements in the commutator subgroup of a given group which is *universal* for certain kinds of extremal problems. For background or proofs, see [1] or [9].

Definition 2.1. Let G be a group. For $g \in [G, G]$ the *commutator length* of g , denoted $\text{cl}(g)$, is the smallest number of commutators in G whose product is equal to g . The *stable commutator length*, denoted $\text{scl}(g)$, is the following limit

$$\text{scl}(g) := \lim_{n \rightarrow \infty} \frac{\text{cl}(g^n)}{n}$$

Note that the function $\text{cl}(g^n)/n$ is subadditive, so the limit in Definition 2.1 exists. Notice further that cl and scl are class functions, and that they are monotone non-increasing under homomorphisms between groups. If we need to emphasize that cl or scl is being calculated in a fixed group, we will use subscripts; hence $\text{cl}_G(g)$ and $\text{scl}_G(g)$.

Remark 2.2. We sometimes extend cl and scl to all of G by defining $\text{cl}(g) = \infty$ if g is not in $[G, G]$, and replacing \lim by \liminf in the definition of scl . Notice that $\text{scl}(g) < \infty$ if and only if some power of g is in $[G, G]$.

The functions cl and scl can be interpreted geometrically. Let X be a connected CW complex with $\pi_1(X) = G$, and let γ be a loop in X whose free homotopy class represents the conjugacy class of g . Then $\text{cl}(g) \leq n$ if and only if there exists an orientable surface S of genus n with one boundary component, and a map $f : S \rightarrow X$ taking ∂S to the free homotopy class of γ .

Remark 2.3. In order to be able to speak interchangeably about loops γ in spaces X as above and their images, we assume in the sequel that all spaces X are such that every free homotopy class of loop can be realized by an embedded circle. This can be achieved, for an arbitrary homotopy type of CW complex X , by multiplying by a sufficiently high dimensional cube.

Genus is not multiplicative under finite covers, but Euler characteristic is. So when we stabilize cl , the relevant geometric quantity to keep track of is derived from Euler characteristic.

Notation 2.4. Let S be a compact, connected, oriented surface. Then set

$$\chi^-(S) = \min(0, \chi(S))$$

Extend χ^- additively to compact, oriented (but not necessarily connected) surfaces S , so that

$$\chi^-(S) = \sum_i \chi^-(S_i)$$

where S_i ranges over the components of S (cf. [25]).

Notation 2.5. Let S be a compact, connected, oriented surface. Let X be a topological space, and $\gamma : S^1 \rightarrow X$ a continuous loop. Let $f : S \rightarrow X$ be such that there is a commutative diagram

$$\begin{array}{ccc} \partial S & \xrightarrow{i} & S \\ \partial f \downarrow & & \downarrow f \\ S^1 & \xrightarrow{\gamma} & X \end{array}$$

where $i : \partial S \rightarrow S$ is the inclusion map, and define $n(S)$ by the identity $\partial f_*[\partial S] = n(S)[S^1]$ in H_1 . If the (oriented) components of ∂S are denoted ∂_i , then $n(S)$ is the sum of the degrees of the maps $\partial f : \partial_i \rightarrow S^1$. Informally, $n(S)$ is the degree with which ∂S wraps around the loop γ .

If $n(S)$ is nonzero, one says that the surface S *rationally bounds* γ . Strictly speaking, $n(S)$ depends on f and not just on S , but we suppress this in our notation.

With these definitions, one can give a geometric interpretation of scl .

Lemma 2.6. *Let X be a connected CW complex with $\pi_1(X) = G$ and let γ be a loop in X in the free homotopy class corresponding to the conjugacy class of g . Then*

$$\text{scl}(g) = \inf_S \frac{-\chi^-(S)}{2|n(S)|}$$

where the infimum is taken over all maps $f : S \rightarrow X$ wrapping ∂S around γ with any degree $n(S)$.

Proof. An inequality in one direction can be obtained by restricting the class of admissible S to those that are connected with exactly one boundary component. To obtain the inequality in the other direction, first observe that components without boundary can be thrown away without increasing $-\chi^-$. Passing to a cover multiplies both $-\chi^-$ and n by the same factor. Moreover an orientable surface with p boundary components admits a cover (in fact, a cyclic cover) of degree m which also has p boundary components, providing m and $p - 1$ are coprime. After passing to such a cover with m very large, multiple boundary components can be tubed together with 1-handles (whose image in X can be taken to be a point), increasing $-\chi^-$ by a term which is arbitrarily small compared to n , thereby proving the theorem. \square

By changing the orientation on S if necessary, we may always take $n(S)$ to be positive. In the sequel we therefore adhere to the convention that $n(S)$ is positive unless we explicitly say otherwise. On the other hand, even if $n(S)$ is positive, if S has more than one boundary component, some components might map to γ with positive degree, and others with negative degree. Say further that a surface S is *monotone* if the degree of every component $\partial_i \rightarrow \gamma$ is positive. The following lemma shows that for the purposes of computing scl , one can restrict attention to monotone surfaces.

Lemma 2.7. *Let $f : S \rightarrow X$ be a connected surface with $\chi(S) < 0$ that rationally bounds γ . Then there is another surface $f' : S' \rightarrow X$ which is monotone, and satisfies $-\chi^-(S')/2n(S') \leq -\chi^-(S)/2n(S)$.*

Proof. Each component ∂_i of ∂S maps to γ with some degree $n_i \in \mathbb{Z}$, where $\sum_i n_i = n(S)$. If some n_i is zero, the image $f(\partial_i)$ is homotopically trivial in X , so we

may reduce $-\chi^-(S)$ without affecting $n(S)$ by compressing ∂_i . So without loss of generality, assume that no n_i is zero.

Since $\chi(S)$ is negative, there is a finite cover of S with positive genus. If S is a connected surface with positive genus and negative Euler characteristic, there is a connected degree 2 cover of S such that each boundary component in S has exactly two preimages in the cover. Hence, after passing to a finite cover if necessary (which does not affect the ratio of $-\chi^-$ to $n(\cdot)$) we can assume that the boundary components ∂_i of S come in *pairs* with equal degrees n_i .

Let N be the least common multiple of the $|n_i|$. Define a function ϕ from the set of boundary components of S to $\mathbb{Z}/N\mathbb{Z}$ as follows. Divide the set of components into pairs ∂_i, ∂_j for which $n_i = n_j$, and define $\phi(\partial_i) = n_i$ and $\phi(\partial_j) = -n_i$. Then $\sum_i \phi(\partial_i) = 0$, so ϕ extends to a surjective homomorphism from $\pi_1(S)$ to $\mathbb{Z}/N\mathbb{Z}$. If S'' is the cover associated to the kernel, then each component of $\partial S''$ maps to γ with degree $\pm N$. Pairs of components whose degrees have opposite sign can be glued up (which does not affect $-\chi^-$ or $n(\cdot)$) until all remaining components have degrees with the same (positive) sign. The resulting surface S' satisfies the conclusion of the lemma. \square

For more details, see [9], § 2.1.

2.2. Extremal surfaces.

Definition 2.8. A map $f : S \rightarrow X$ rationally bounding γ is *extremal* if S has no disks or closed components, and there is an equality $\text{scl}(g) = -\chi^-(S)/2n(S)$.

Notice that $\text{scl}(g)$ must be rational for an extremal surface to exist.

Lemma 2.9. *An extremal surface is π_1 -injective.*

Proof. Let $f : S \rightarrow X$ be extremal. Suppose there is some essential immersed loop α in S for which $f(\alpha)$ is null-homotopic. Since surface groups are LERF (see [23]) there is a finite cover \hat{S} of S to which α lifts as an embedded loop. Let $\hat{f} : \hat{S} \rightarrow X$ lift the map f (i.e. \hat{f} is the composition of f with the covering projection $\hat{S} \rightarrow S$). Note that $-\chi^-(\hat{S})/2n(\hat{S}) = -\chi^-(S)/2n(S)$.

Let $\hat{\alpha}$ denote an embedded preimage of $\alpha \subset S$ in \hat{S} . Since $\hat{\alpha}$ is nullhomotopic under \hat{f} , we can surger \hat{S} along $\hat{\alpha}$ to produce a surface S' with $-\chi^-(S') < -\chi^-(\hat{S})$ but with $n(S') = n(\hat{S})$. But this contradicts the hypothesis that S is extremal. \square

Lemma 2.7 shows that if there is an extremal surface for γ , there is a monotone extremal surface.

2.3. Quasimorphisms. A brief discussion of quasimorphisms, though not strictly logically necessary for the results of this paper, nevertheless provides some useful context and explains an important connection with the theory of bounded cohomology.

Definition 2.10. Let G be a group. A *quasimorphism* on G is a function $\phi : G \rightarrow \mathbb{R}$ for which there exists some least non-negative constant $D(\phi)$ called the *defect*, so that the following inequality holds

$$|\phi(g) + \phi(h) - \phi(gh)| \leq D(\phi)$$

for all $g, h \in G$. A quasimorphism is *homogeneous* if $\phi(g^n) = n\phi(g)$ for all $g \in G$ and all $n \in \mathbb{Z}$.

In words, a quasimorphism on a group is a homomorphism up to a bounded error. A quasimorphism is a genuine homomorphism if and only if the defect is zero.

The set of quasimorphisms (resp. homogeneous quasimorphisms) on G admits the structure of a real vector space. Denote the vector space of quasimorphisms on G by $\widehat{Q}(G)$, and the vector space of homogeneous quasimorphisms by $Q(G)$.

Proposition 2.11 (Bavard [1], Prop. 3.3.1). *Let G be a group. There is an exact sequence*

$$0 \rightarrow H^1(G; \mathbb{R}) \rightarrow Q(G) \xrightarrow{\delta} H_b^2(G; \mathbb{R}) \rightarrow H^2(G; \mathbb{R})$$

where H_b^* denotes bounded cohomology (with real coefficients), and δ denotes the coboundary on group 1-cochains.

See [14] for an introduction to bounded cohomology. Note that when H^1 and H^2 are finite dimensional (as is the case when G is finitely presented) then δ is Fredholm (with respect to natural Banach norms on Q/H^1 and H_b^2).

There is a kind of duality, called *Bavard duality*, relating commutator length and quasimorphisms. The most concise statement of this duality is the following:

Theorem 2.12 (Bavard's Duality Theorem [1], p. 111). *Let G be a group. For any $g \in [G, G]$ there is an equality*

$$\text{scl}(g) = \frac{1}{2} \sup_{\phi \in Q(G)} \frac{\phi(g)}{D(\phi)}$$

Note that one should restrict attention to $\phi \in Q(G) - H^1(G)$ since if ϕ is a homomorphism, then both $\phi(g) = 0$ for $g \in [G, G]$, and $D(\phi) = 0$.

2.4. Stable commutator length as a norm. The functions cl and scl can be extended to finite sums as follows.

Definition 2.13. Let G be a group. Let g_1, \dots, g_m be elements in G (not necessarily distinct). Define

$$\text{cl}(g_1 + g_2 + \dots + g_m) = \inf_{h_1, \dots, h_{m-1} \in G} \text{cl}(g_1 h_1 g_2 h_1^{-1} h_2 g_3 h_2^{-1} \dots h_{m-1} g_m h_{m-1}^{-1})$$

and define

$$\text{scl}(g_1 + g_2 + \dots + g_m) = \lim_{n \rightarrow \infty} \frac{\text{cl}(g_1^n + \dots + g_m^n)}{n}$$

Note that cl and scl depend only on the individual conjugacy classes of the summands, and are commutative in their arguments. Geometrically, if X is a CW complex with $\pi_1(X) = G$ and $\gamma_1, \dots, \gamma_m$ are loops representing the conjugacy classes of g_1, \dots, g_m respectively, then $\text{cl}(\sum g_i)$ is the smallest genus surface S with m boundary components ∂_i for which there is a map $f : S \rightarrow X$ wrapping each ∂_i around γ_i . It is worth remarking that the function $\text{cl}_n := \text{cl}(\sum g_i^n)$ is not subadditive, but that the “corrected” function $\text{cl}_n + (m - 1)$ is subadditive, and therefore the limit exists in Definition 2.13, providing cl is not infinite.

Notation 2.14. Let S be a compact, connected, oriented surface. Let X be a topological space, and $\gamma_i : S^1 \rightarrow X$ for $1 \leq i \leq m$ continuous loops. Let $f : S \rightarrow X$

be such that there is a commutative diagram

$$\begin{array}{ccc} \partial S & \xrightarrow{i} & S \\ \partial f \downarrow & & \downarrow f \\ \coprod_i S^1 & \xrightarrow{\coprod_i \gamma_i} & X \end{array}$$

where $i : \partial S \rightarrow S$ is the inclusion map. Suppose there is an integer $n(S)$ so that $\partial f_*[\partial S] = n(S)[\coprod_i S^1]$ in H_1 . Then say $f : S \rightarrow X$ is *admissible*. Informally, $n(S)$ is the common degree with which ∂S wraps around each loop γ_i .

With this notation, the generalization of Lemma 2.6 to arbitrary sums is as follows:

Lemma 2.15. *Let X be a connected CW complex with $\pi_1(X) = G$. Further, let $\gamma_1, \dots, \gamma_m$ be loops in X in free homotopy classes corresponding to conjugacy classes g_1, \dots, g_m . Then*

$$\text{scl}(\sum_i g_i) = \inf_S \frac{-\chi^-(S)}{2|n(S)|}$$

where the infimum is taken over all admissible maps $f : S \rightarrow X$ wrapping ∂S around each γ_i with degree $n(S)$.

The proof is almost identical to that of Lemma 2.6. Moreover, one may restrict attention to monotone admissible maps, by the argument of Lemma 2.7. For details, see [9], § 2.6.1.

The function scl as above can be extended to integral group 1-chains. From Lemma 2.15 follow equalities

$$\text{scl}(g + g^{-1} + \sum g_i) = \text{scl}(\sum g_i)$$

and

$$\text{scl}(g^n + \sum g_i) = \text{scl}(\underbrace{g + \dots + g}_n + \sum g_i)$$

valid for any g, g_i and any non-negative integer n . Hence one may define

$$\text{scl}(\sum n_i g_i) := \text{scl}(\sum g_i^{n_i})$$

for any integers n_i (not necessarily non-negative) and elements $g_i \in G$ and observe that the result is well-defined on integral group 1-chains, and is subadditive under addition of chains. Consequently, scl can be extended to rational chains by linearity, and to real chains by continuity. See [9], § 2.6.1.

Denote the vector space of real (group) 1-chains on G by $C_1(G)$ and 1-boundaries by $B_1(G)$ (or just C_1 and B_1 if G is understood). The expression $\text{scl}(\sum t_i g_i)$ is finite if and only if $\sum t_i g_i \in B_1$. Bavard duality holds in the broader context of arbitrary real 1-boundaries, and with essentially the same proof. The statement is:

Theorem 2.16 (Generalized Bavard Duality [9] § 2.6.2). *Let G be a group. For any finite set of elements $g_i \in G$ and numbers $t_i \in \mathbb{R}$ for which $\sum t_i g_i \in B_1$ there is an equality*

$$\text{scl}(\sum t_i g_i) = \frac{1}{2} \sup_{\phi \in Q(G)} \frac{\sum t_i \phi(g_i)}{D(\phi)}$$

Any homogeneous quasimorphism vanishes identically on any chain of the form $g^n - ng$ or $g - hgh^{-1}$ for $g, h \in G$ and $n \in \mathbb{Z}$. Define H to be the subspace of B_1 spanned by such chains. Then scl descends to a pseudo-norm on $B_1(G)/H$ (hereafter denoted B_1^H). When G is hyperbolic, scl is a *norm* on B_1^H ; this is Corollary 3.57 from [9], restating work of [10] (this fact is logically superfluous for the results of this paper).

In fact, the reader can take Lemma 2.6 and Lemma 2.15 as the *definition* of scl . This is the point of view we shall take in the sequel.

3. FREE GROUPS

This section contains the proof of the Rationality Theorem for free groups. The main result is that scl is a piecewise rational linear function on $B_1^H(F)$, for F a free group. A similar statement holds for some groups derived in simple ways from free groups. Throughout this section, we use Lemma 2.6 and Lemma 2.15 as an operational definition of scl .

3.1. Handlebodies and arcs. In this section, for convenience, we use some language and basic facts from elementary 3-manifold topology; for a reference, see [16]. In the sequel, let F denote a free group of some fixed rank and let H denote a handlebody of genus equal to $\text{rank}(F)$. As in Figure 1 (illustrating the case of $\text{rank}(F) = 4$), we consider a system of compressing disks D_i which decompose H into $\text{rank}(F) - 1$ components, each of which retracts down to one of $\text{rank}(F) - 1$ compressing disks E_j . Denote the union of the D_i by \mathcal{D} and the union of the E_j by \mathcal{E} .

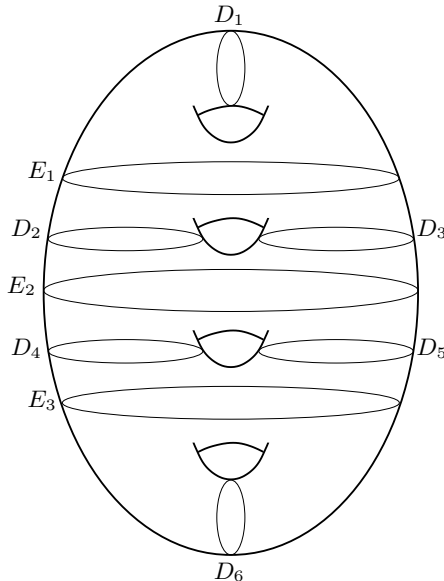


FIGURE 1. The decomposing disks E_i and D_j for $g = 4$

Given a conjugacy class in F , we construct a representative loop in the corresponding free homotopy class in H of a simple kind. Such a representative will be made up of certain kinds of arcs, which we call *horizontal* and *vertical*, and which are defined as follows.

Definition 3.1. A *horizontal* arc is an embedded arc $\alpha : I \rightarrow H$ whose image is contained in some E_i . A *vertical* arc is an arc $\alpha : I \rightarrow H$ which is properly embedded in the complement of \mathcal{E} , and which intersects some D_j transversely in one point.

Note that any two horizontal arcs with the same endpoints are homotopic rel. endpoints through horizontal arcs. Moreover, any two vertical arcs whose endpoints are contained in the same E_i are properly homotopic through vertical arcs; call arcs which differ from each other by such homotopies *equivalent*.

Remark 3.2. If one does not want the psychological convenience of working in a manifold, one can substitute in place of H a union of $\text{rank}(F)$ solid tori H_i , each with a marked disk E_i in their boundary, and glue the tori up by identifying the E_i with a single disk E by homeomorphisms. The resulting space is a manifold away from the disk E . In the case $\text{rank}(F) = 2$, the two approaches are equivalent.

Dual to the system \mathcal{D} of compressing disks there is a graph Γ with one vertex for each component of $H - \mathcal{D}$ and one edge for each D_i in \mathcal{D} . There is an isomorphism $\pi_1(\Gamma) \cong \pi_1(H) \cong F$. The universal cover $\tilde{\Gamma}$ of Γ is a tree (for an introduction to trees in geometric group theory, see [24], especially Chapter 1). Every element in F acts on the tree $\tilde{\Gamma}$ with a unique axis; this axis covers a closed loop in Γ . Each arc in Γ corresponds to a unique equivalence class of vertical arc in H . So to each conjugacy class of element $g \in F$ is associated a (cyclically ordered) sequence of (equivalence classes of) vertical arcs in H . If two consecutive vertical arcs are on opposite sides of some E_i in \mathcal{E} , they can be homotoped (rel. endpoints) until they have a common endpoint in E_i , and their union is transverse to E_i at that point. If two consecutive vertical arcs are on the same side of some E_i in \mathcal{E} , their endpoints in E_i can be joined by a horizontal arc in E_i . In this way, a conjugacy class in F determines a loop γ in H , unique up to equivalence, made up of vertical and horizontal arcs. Say that such a γ is in *bridge position*.

If g_1, g_2, \dots, g_m is a family of elements in F , then a family of loops $\gamma_1, \dots, \gamma_m$ in H representing the conjugacy classes of the g_i is in bridge position if each loop individually is in bridge position.

See Figure 2 for an example of a loop in bridge position. Since each horizontal or vertical arc has distinct endpoints, a circle in bridge position decomposes into at least two arcs. We may (and do) assume without loss of generality that any family of circles in bridge position is actually embedded in H . However, it is very important to note that the *isotopy class* of the loop is *not* important; all that matters is its *homotopy class*.

Remark 3.3. A system \mathcal{D} and \mathcal{E} of compressing disks for a handlebody determines a generating set for F as a groupoid, whose generators are equivalence classes of vertical arcs. If we use in place of a handlebody the space described in Remark 3.2, vertical arcs correspond to generators for F as a group.

3.2. Polygons and rectangles. Now, let $g \in F$ be a conjugacy class in $[F, F]$, and let γ be a loop in bridge position representing g . Let $Z \subset \mathcal{E}$ be the union of the endpoints of the horizontal and vertical arcs in γ (remember that for convenience we have assumed that γ is embedded). Note that Z is a finite set. Let $f : S \rightarrow H$ be a map of a surface whose (possibly multiple) boundary components wrap some number of times around γ . By Lemma 2.7, we may assume that f is monotone, so

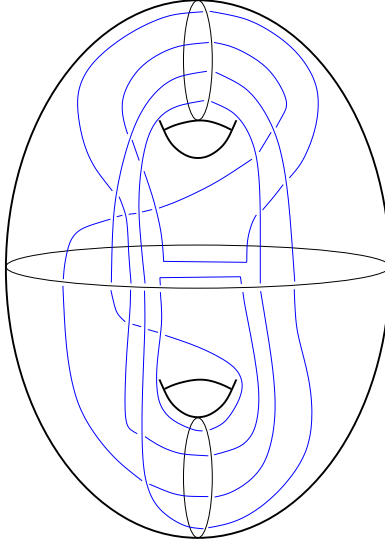


FIGURE 2. A loop in bridge position representing the element $ababa^{-2}b^{-2}$ in F_2 . There are eight vertical arcs (one for each letter) and two horizontal arcs (one for each “double letter”). This loop happens to be embedded in H , but the isotopy class of the loop is not significant, just its homotopy class.

that each component of ∂S maps over γ with positive degree. This is not strictly necessary for what follows, but it simplifies some arguments. We will gradually adjust f and S , never changing $n(S)$ or increasing $-\chi^-$, until the end result is built up from a finite number of simple “pieces”.

Since by hypothesis f is monotone, first adjust f by a homotopy so that the restriction of f to each boundary component is an orientation-preserving covering map $\partial_i \rightarrow \gamma$. Next put f into general position (rel. ∂S) with respect to the disks \mathcal{D} .

Since f is in general position, and since γ is transverse to \mathcal{D} , the preimage $f^{-1}(\mathcal{D})$ is a union of disjoint embedded loops and proper arcs. Since f restricted to ∂S is an immersion, every arc of $f^{-1}(\mathcal{D})$ is essential. Let α be a component of $f^{-1}(\mathcal{D})$ in S . If α is an (innermost) inessential loop, then α can be pushed off \mathcal{D} by a homotopy, reducing the number of components of $f^{-1}(\mathcal{D})$. If α is an essential loop, then S can be compressed along α , and the compressing disks can be mapped to the component of \mathcal{D} containing $f(\alpha)$. This does not change $n(S)$ but reduces $-\chi^-$. After finitely many such compressions and homotopies, we can assume that $f^{-1}(\mathcal{D})$ is a union of disjoint embedded proper essential arcs. In short, we have proved the following “preparation lemma”:

Lemma 3.4. *Let $f : S \rightarrow H$ monotone be given. Then after possibly replacing f, S with f', S' satisfying $-\chi^-(S') < -\chi^-(S)$ and $n(S') = n(S)$, we can assume that $f^{-1}(\mathcal{D})$ is a union of disjoint embedded proper essential arcs.*

Note that distinct components of $f^{-1}(\mathcal{D})$ might be (and typically will be) parallel in S , especially for complicated γ . Let \mathcal{R} be a regular neighborhood of $f^{-1}(\mathcal{D})$ in S , so that \mathcal{R} consists of a union of disjoint embedded proper essential rectangles.

By general position we can take \mathcal{R} to be equal to the inverse image (under f^{-1}) of an open tubular neighborhood $N(\mathcal{D})$ of \mathcal{D} . Now, there is a deformation retraction of pairs $H - N(\mathcal{D}), \gamma \cap (H - N(\mathcal{D}))$ to $\mathcal{E}, \gamma \cap \mathcal{E}$. This deformation retraction can be extended to a map $r : H \rightarrow H$ which restricts to a homotopy equivalence of pairs from $N(\mathcal{D}), \gamma \cap N(\mathcal{D})$ to $H - \mathcal{E}, \gamma \cap (H - \mathcal{E})$. So after composing f with such a retraction r we get a new map, which by abuse of notation we also denote by f , homotopic to the old map f , such that $\mathcal{R} = f^{-1}(H - \mathcal{E})$. Each component β of $\partial S - \mathcal{R}$ either maps by f to a point in Z , or to a horizontal arc in γ . In the first case, we collapse β to a point in S by a homotopy equivalence, so we assume without loss of generality that every arc of $\partial S - \mathcal{R}$ is a horizontal arc. See Figure 3.

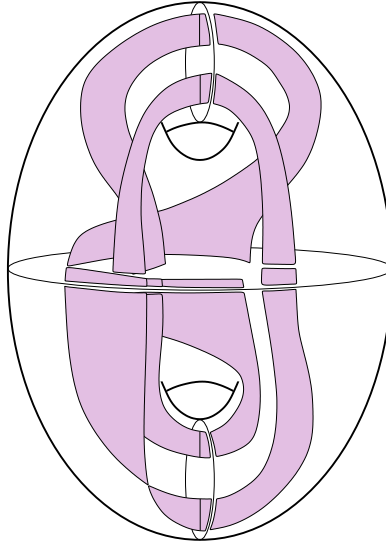


FIGURE 3. After retracting $H - N(\mathcal{D})$ to E , the preimage $f^{-1}(H - E)$ consists of a union \mathcal{R} of rectangles.

Let P be a component of $S - \mathcal{R}$. Then f maps P to some component E_i of E . If P is not a disk, then it contains an essential simple loop γ . Since E_i is a disk, f maps γ to a homotopically trivial loop in E_i , so we can compress S along γ , mapping the compressing disks to E_i , to get a new surface S' with $-\chi^-(S') < -\chi^-(S)$ and $n(S') = n(S)$. So without loss of generality we can assume that each component P is a disk. In fact, P inherits the structure of a polygon, whose edges are arcs of the boundary of components of \mathcal{R} , and horizontal arcs. The vertices of P map by f to Z .

Lemma 3.5. *With notation as above, after possibly replacing f, S with f', S' satisfying $-\chi^-(S') < -\chi^-(S)$ and $n(S') = n(S)$ we can assume that for each polygon P of $S - \mathcal{R}$ the image of the vertices of P under f are distinct elements of Z .*

Proof. Suppose that P is a polygon mapping under f to the disk E , and let u, v be distinct vertices of P mapping to the same point $y \in Z \cap E$. Let β be an embedded proper arc in P from u to v . Let S'' be obtained from S by identifying u to v , then cut S'' along the loop β , and glue in two disks. The restriction of f to $S'' - \beta$

extends to these disks, since E is contractible. The result is a new surface S' as in the statement of the Lemma. \square

Remark 3.6. The argument of Lemma 3.5 may be summarized by saying that S' is obtained from S by first adding an oriented 1-handle from u to v which maps trivially under f , and then compressing the trivial embedded loop that runs over the core of this 1-handle. Adding a 1-handle increases $-\chi^-$ by 1, but doing a compression reduces it by 2. These operations are uniquely defined up to homotopy, and keep the surface oriented and the map monotone. Notice that S and S' might have different numbers of boundary components. Call the result of this whole operation a *boundary compression*.

3.3. Simple branched surfaces. The core of each component of \mathcal{R} runs between two vertical arcs of γ , after mapping by f . There are only finitely many combinatorial types of such rectangles; an upper bound is the number of pairs of vertical arcs in γ which intersect the same component of \mathcal{D} . In fact, since S is monotone, the two (oriented) arcs of γ in the boundary of each rectangle must run over a handle of H in opposite directions. Similarly, there are only finitely many polygon types P , since each is determined by a cyclically ordered subset of $Z \cap E_i$ for some E_i .

It is convenient to introduce the language of branched surfaces in what follows. We give cursory definitions below to standardize terminology, but the reader who is not familiar with branched surface should consult e.g. [4] Chapter 6, § 6.3, or [19] § 1 for precise definitions.

A *branched surface* is a finite, smooth 2-complex obtained from a finite collection of smooth surfaces by gluing compact subsurfaces. The set of non-manifold points of a branched surface B is called the *branch locus*, and denoted $C(B)$. A branched surface in which the branch locus is a 1-manifold is called a *simple branched surface*. In this paper we are only interested in simple branched surfaces. One can also consider branched surfaces with boundary; the boundary of a branched surface is a *train-track* (see e.g. [20]). The components of $B - C(B)$ are called the *sectors* of the branched surface. A branched surface is oriented if the sectors can be oriented in such a way that the orientations are compatible along $C(B)$. In a simple branched surface, several distinct sectors locally bound a component of the branch locus from either side; see Figure 4 for an example of the local model. A *weight* is a linear

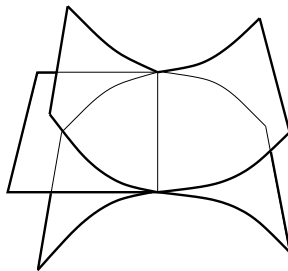


FIGURE 4. An arc of the branch locus with two sectors on one side and three sectors on the other

function w from the sectors of B to \mathbb{R} satisfying the *gluing conditions*: along each component of $C(B)$, the weights of the sectors on each side sum to the same value.

A branched surface B has a well-defined tangent bundle, even along the branch locus, so it makes sense to say that a map from a surface into B is an immersion. A proper immersion $f : S \rightarrow B$ from an oriented surface to an oriented branched surface B is said to be a *carrying map*, and S is said to be *carried* by B . A carrying map determines a weight w , where $w(\sigma)$ is the cardinality of $f^{-1}(p)$ for p a point in the interior of a sector σ . Given a non-negative integral weight w , we say that a carrying map $f : S \rightarrow B$ *realizes* w if the weight associated to S is w . Note that S is not uniquely determined by its weight in general.

For an abstract branched surface, not every non-negative integral weight is realized, but for a *simple* branched surface, every non-negative integral weight is realized. This is a somewhat subtle point, so we make a few clarifying remarks. Let w be a weight. An edge e of $C(B)$ has sectors $\sigma_1, \dots, \sigma_m$ on one side, and $\sigma_{m+1}, \dots, \sigma_n$ on the other. Take $w(\sigma_i)$ copies of σ_i for each i . The gluing condition along e is the equality

$$\sum_{i=1}^m w(\sigma_i) = \sum_{i=m+1}^n w(\sigma_i)$$

Hence there is a bijection between the set of copies of the $\sigma_1, \dots, \sigma_m$ and the set of copies of the $\sigma_{m+1}, \dots, \sigma_n$. After choosing such a bijection, these copies can be glued up along copies of e to build the surface S near e . For a general branched surface, there is a holonomy problem at the triple points of $C(B)$: around each triple point, the holonomy must be trivial, or else the surface constructed will map to B with branch points. But if B is simple, there are no such triple points, and the holonomy problem goes away. This concludes our summary of the theory of branched surfaces.

Remark 3.7. If B is an *embedded* branched surface (possibly with triple points) in a 3-manifold M , the local transverse order structure canonically solves the holonomy problem: associated to each weight there is a unique *embedded* surface contained in a tubular neighborhood of B . The requirement that the surface be embedded makes the bijections along edges canonical, and therefore the holonomy around a triple point is necessarily trivial. A similar phenomenon occurs in normal surface theory: a vector of weights satisfying the gluing equalities and inequalities determines a unique embedded normal surface. When one tries to do immersed normal surface theory in 3-manifolds, the holonomy problem reasserts itself and the situation becomes very tricky; see [22] for a discussion of some of the phenomena which arise in the theory of immersed normal surfaces.

We construct a branched surface B as follows. B is obtained by taking a disjoint copy of each combinatorial type of (oriented) marked polygon P , and a copy of each combinatorial type of (oriented) rectangle R as above, and gluing them along their common oriented edges. There is a unique way to choose the smooth structure along $C(B)$ compatibly with the orientations. Notice that each marked polygon is glued to exactly one rectangle along each non-horizontal edge, but each rectangle is typically glued to several polygons along each of two edges. Hence a weight is determined by its values on polygon sectors (cf. quadrilateral co-ordinates in normal surface theory). The resulting branched surface is special in a few ways, of which we take note:

- (1) It is *oriented*, since it is obtained by gluing oriented pieces in a manner compatible with their orientations.
- (2) The branch locus consists of a finite union of embedded arcs along which a rectangle is glued to several possible polygons. In particular, the branch locus contains no triple points and is *simple*.
- (3) The branched surface has boundary which is an oriented train-track. Every surface S carried by B has boundary ∂S which is carried by ∂B .
- (4) The branched surface admits a tautological immersion $\iota : B \rightarrow H$, by choosing a map for each rectangle and polygon type and gluing these maps together along the edges. The restriction of ι to each train-track component of ∂B is an oriented immersion to γ .

For each polygon P let $s(P)$ denote the number of edges, and $h(P)$ the number of horizontal edges. Note that $s(P) \geq 2$ and $h(P) \leq s(P)/2$, since a pair of adjacent edges of a polygon cannot both be horizontal. Furthermore, if $s(P) = 2$ then $h(P) = 0$. Let $g : S \rightarrow H$ be decomposed into polygons and rectangles. Each rectangle contributes 0 to $\chi(S)$, and each polygon P contributes

$$\chi(P) = \frac{-(s(P) - 2 - h(P))}{2}$$

This can be seen by giving each rectangle and polygon a singular foliation tangent to ∂B and transverse to $C(B)$, and using the Hopf-Poincaré formula.

In particular, the contribution of each polygon to $\chi(S)$ is *nonpositive*, and therefore $-\chi^-(S)$ is a *linear function* of the number and kinds of rectangles and polygons that make it up.

In summary, the results of the previous sections show that for any monotone $f : S \rightarrow H$, after possibly replacing S by a new surface with smaller $-\chi^-$ and the same $n(S)$, we can homotope f so that it factors through a carrying map to B , and determines a non-negative integral weight. Conversely, since B is simple, every non-negative integral weight on B corresponds to a surface S as above (possibly not unique), and $-\chi^-$ is a linear function of the values of w on the branches of B . Notice that $-\chi^-$ depends only on w , though the topology of S might not.

Let W denote the (finite dimensional) real vector space of weights on B , and W^+ the subspace of nonnegative weights. Moreover, let $W(\mathbb{Q})$ (resp. $W^+(\mathbb{Q})$) denote the subset of weights with rational (resp. non-negative rational) coefficients. Then $-\chi^-$ is a linear function on W which is non-negative on W^+ , and takes rational values on $W(\mathbb{Q})$. We abbreviate this by saying that $-\chi^-$ is a *rational linear function*.

Let V be the subspace of W spanned by W^+ (note that this subspace is not necessarily equal to W). Then V is a rational subspace of W , since it is spanned by finitely many rational vectors. There is a rational linear function $\partial : V \rightarrow \mathbb{R}$ defined as follows. Given a positive integral weight w , let S be a surface carried by B associated to w . Then define

$$\partial(w) = n(S)$$

and extend by linearity to V . Notice that $\partial(w)$ does not depend on the choice of S , but only on the induced weight on the train-track ∂B .

The inverse $\partial^{-1}(1) \cap W^+$ is a rational polyhedron. Moreover, by construction, there is an equality

$$\text{scl}(g) = \inf_{w \in \partial^{-1}(1) \cap W^+} \frac{-\chi^-(w)}{2}$$

Since $-\chi^-$ is non-negative on W^+ , this infimum is *realized*, and the set of points which realize the infimum is itself a rational polyhedron. If w is an integral weight in the projective class of an element of this polyhedron and S is a surface carried by B with weight w , then S is an extremal surface for g .

From this discussion we can conclude the following:

- (1) $\text{scl}(g) \in \mathbb{Q}$ for every $g \in [F, F]$
- (2) An extremal surface exists for every g
- (3) There is an algorithm to calculate scl and to construct all monotone extremal surfaces for every $g \in [F, F]$

Remark 3.8. If $f : S \rightarrow H$ is extremal, and there is an essential embedded loop α in S such that $f(\alpha)$ is freely homotopic to a power of γ , we may cut open S along α to produce a new extremal (but not necessarily monotone) surface.

3.4. Polyhedral norm. In fact, very little is required to extend the results of the last few sections to finite dimensional vector spaces of $B_1(F)$. Let $g_1, g_2, \dots, g_m \in F$ be represented by loops $\gamma_1, \gamma_2, \dots, \gamma_m$ in bridge position in H . Denote $\gamma = \cup_i \gamma_i$. Fix an orientation on each γ_i . A monotone surface $f : S \rightarrow H$ whose boundary wraps some positive number of times around the various γ_i can be compressed, boundary compressed and homotoped until it is composed of a union of rectangles and polygons carried by a fixed simple branched surface B as above. There is a rational linear map $\partial : V \rightarrow H_1(\bigcup_i \gamma_i; \mathbb{R}) \cong \mathbb{R}^m$.

Let K be the kernel of the inclusion map $H_1(\bigcup_i \gamma_i; \mathbb{R}) \rightarrow H_1(H; \mathbb{R})$. Let K^+ be the intersection of K with the orthant spanned by non-negative combinations of the $[\gamma_i]$ in $H_1(\bigcup_i \gamma_i; \mathbb{R})$. Given $k \in K^+$ corresponding to a collection of non-negative weights on the g_i , we have an equality

$$\text{scl}(k) = \inf_{w \in \partial^{-1}(k) \cap W^+} \frac{-\chi^-(w)}{2}$$

and therefore scl is a piecewise rational linear function on K^+ . There are finitely many orthants of this kind, corresponding to choices of orientation on each γ_i , so scl is piecewise rational linear on K .

Putting this together proves our main result:

Rationality Theorem. *Let F be a free group.*

- (1) $\text{scl}(g) \in \mathbb{Q}$ for all $g \in [F, F]$.
- (2) Every $g \in [F, F]$ rationally bounds an extremal surface (in fact, every rational chain C in B_1^H rationally bounds an extremal surface)
- (3) The function scl is piecewise rational linear on B_1^H .
- (4) There is an algorithm to calculate scl on any finite dimensional rational subspace of B_1^H .

Note that bullet (1) is a special case of bullet (3).

3.5. Other groups.

Definition 3.9. Say that a group G is PQL (pronounced “pickle”) if scl is piecewise rational linear on $B_1^H(G)$.

The PQL property is inherited by supergroups of finite index. This follows in a straightforward way from the following Lemma, which relates scl in groups and in finite index subgroups.

Lemma 3.10. *Let G be a group, and H a subgroup of G of finite index. Let X be a CW complex with $\pi_1(X) = G$, and let \hat{X} be a covering space with $\pi_1(\hat{X}) = H$. Let g_1, \dots, g_m be elements in G , and for each i , let γ_i be a loop in X representing the conjugacy class of g_i . Let β_1, \dots, β_l be the preimages of the γ_i in \hat{X} , and h_1, \dots, h_l the corresponding conjugacy classes in H . Then*

$$\text{scl}_H(\sum h_i) = |G : H| \cdot \text{scl}_G(\sum g_i)$$

Proof. If $f : S \rightarrow X$ is a map of a surface whose boundary maps to the γ_i , then S admits a finite index cover \hat{S} such that f lifts to $\hat{f} : \hat{S} \rightarrow \hat{X}$. Conversely, given $f : S \rightarrow \hat{X}$ with boundary mapping to the β_i , the composition of f with the covering projection $\hat{X} \rightarrow X$ takes ∂S to γ . Now apply Lemma 2.6 and Lemma 2.15. \square

Theorem 3.11. *Let M be a non-compact Seifert-fibered 3-manifold. Then $\pi_1(M)$ has the PQL property.*

Proof. Since M is noncompact, there is a finite index subgroup H of $\pi_1(M)$ of the form $H = \mathbb{Z} \oplus F$ where F is free. Since \mathbb{Z} is amenable, a theorem of Bouarich (see [2] or [9], § 2.4) implies that $H_b^2(F) \rightarrow H_b^2(H)$ is an isomorphism, and $Q(H) = Q(F) \oplus H^1(\mathbb{Z})$. Hence H is PQL by Theorem 2.16. But then $\pi_1(M)$ is PQL by Lemma 3.10. \square

Example 3.12. The braid group B_3 is isomorphic to $\pi_1(S^3 - K)$ where K is the trefoil knot (of either handedness). This group is a central \mathbb{Z} extension of $\mathbb{Z}/2\mathbb{Z} * \mathbb{Z}/3\mathbb{Z}$, which admits a free subgroup of finite index. Hence B_3 has the PQL property.

Question 3.13. *Does every 3-manifold group have the PQL property?*

4. COMPUTING STABLE COMMUTATOR LENGTH

In this section we discuss the implementation of the algorithm implicit in § 3, and study an explicit example. In order to describe the algorithm in a uniform way for free groups of any rank, it is convenient to use the formalism described in Remark 3.2, in which vertical arcs in Γ correspond to elements in a free generating set for F . If one is more comfortable working with systems of compressing disks in a handlebody, one must work with *groupoid* generators for F associated to the system of vertical arcs in a splitting. In the case of a free group of rank 2, both formalisms agree; the cautious reader may prefer to stick to this case in what follows.

Fix a free symmetric generating set S for a free group F , and let $C = w_1 + w_2 + \dots + w_m$ be an integral chain in $B_1^H(F)$, expressed as a formal sum of cyclically reduced words in the generating set. We denote the generators of F by a, b, c, \dots and their inverses by A, B, C, \dots .

Definition 4.1. A *letter* is a specific character in a specific word; its *value* is the element of the generating set that it represents. An *arc* is an ordered pair (ℓ_1, ℓ_2) of letters whose values are inverse in F . A *polygon* is a cyclically ordered list of distinct arcs $\alpha_0, \alpha_1, \dots, \alpha_{m-1}$ so that for each i , the last letter of α_i immediately precedes the first letter of α_{i+1} in some (cyclic) word w_j (indices i and $i+1$ are taken mod m). The *length* of a polygon is the number of arcs appearing in the list.

In the language of § 3, each oriented rectangle has four edges, two of which (the “free edges”) correspond to vertical edges in Γ , and two of which (the “glued edges”) are glued up to some polygon. Each vertical edge in Γ corresponds to a letter in some w_i . The two glued edges of a rectangle determine two arcs α and α' consisting of the same pair of letters in opposite orders; i.e. if $\alpha = (\ell_1, \ell_2)$ then $\alpha' = (\ell_2, \ell_1)$.

A polygon (in the sense of § 3) has edges that are glued up to edges of rectangles, and edges on horizontal edges of Γ . Horizontal edges are of purely psychological value; they do not contribute to χ . So to each such polygon there is associated a polygon (in the sense of Definition 4.1) consisting of a cyclically ordered list of arcs. The condition that the arcs making up a polygon are distinct is the analog of the condition that each polygon (in the sense of § 3) has vertices mapping to distinct points of Z .

Let P denote the real vector space spanned by the set of polygons. Let A denote the vector space spanned by the set of arcs modulo the relation that $(\ell_1, \ell_2) = -(\ell_2, \ell_1)$. There is a linear map $\partial : P \rightarrow A$ sending a polygon to the formal sum of the arcs that make it up. Let W denote the kernel of ∂ , and W^+ the cone of non-negative vectors in W . The relations $(\ell_1, \ell_2) = -(\ell_2, \ell_1)$ and the condition $\partial = 0$ express the gluing equations for B in terms of weights on polygon sectors. The spaces W and W^+ are naturally isomorphic to the spaces with the same names described in § 3. A polygon of length m contributes $(m-2)/4$ to $-\chi^-/2$. The condition that a formal sum of polygons has boundary equal to the chain C is a further list of linear conditions, one for each word in C . Minimizing the objective function $-\chi^-/2$ is a linear programming problem, which can be solved in a number of ways.

Example 4.2. Let $w = ababABaBAbAB$ (remember that $A = a^{-1}$ and $B = b^{-1}$). There are 36 arcs, one for each ordered pair of letters in w with opposite values. Labeling the letters of w by integers from 0 to 11, and using the shorthand $X = 10$ and $Y = 11$, these arcs are

$$\begin{aligned} &04, 40, 08, 80, 0X, X0, 15, 51, 17, 71, 1Y, Y1, 24, 42, 28, 82, 2X, X2, \\ &35, 53, 37, 73, 3Y, Y3, 46, 64, 59, 95, 68, 86, 6X, X6, 79, 97, 9Y, Y9 \end{aligned}$$

There are 625 polygons, 43 of length 4 or less (since the letters of w alternate between one of a^\pm and one of b^\pm , every polygon has even length), including

$$(04, 51, 28, 9Y), (04, 53, 40, 1Y), \dots, (X6, 79)$$

Let $P = \mathbb{R}^{625}$ denote the vector space of formal linear combinations of polygons, and let p_i for $0 \leq i \leq 624$ denote the components of a vector in P . For each i , let l_i denote the length of polygon i . Compatibility of gluing along rectangles (i.e. $\partial = 0$ above) imposes one equation of the form $\sum p_k = \sum p_l$ for each pair of arcs ij, ji where the polygons of type k are those that contain the arc ij , and polygons of type l are those that contain the arc ji (note that a polygon type might contain

both ij and ji or neither). There are half as many equations of this kind as arcs, hence 18 equations.

Restricting to geometrically sensible answers imposes the conditions $p_i \geq 0$. The condition that the boundary of a (formal) surface corresponding to a weight represents $[\gamma]$ in homology is the equation

$$\sum_i l_i p_i = |w| = 12$$

Subject to this list of constraints, which determine a compact convex polyhedron in \mathbb{R}^{625} , we minimize the objective function

$$\frac{-\chi^-}{2} = \sum_i \frac{(l_i - 2)p_i}{4}$$

This linear programming problem can be solved using exact arithmetic by the GNU package `glpsol` [18] or Masashi Kiyomi's program `exlp` [17] using Dantzig's simplex method (see [11]), and gives the answer $\text{scl}(w) = 5/6$.

An extremal solution found by the simplex method is always a vertex, which can be projectively represented by an extremal surface. One such extremal surface found by this method is determined by the identity

$$\begin{aligned} &[abaB, ABAbabABaBABAbabABB] \cdot [ABAb, BabAbABABba] \\ &\cdot [BabABababA, aaBAAb] = a(baBABAbabA)^3 A \end{aligned}$$

exhibiting the cube of $baBABAbabA$ (i.e. the inverse of a cyclic conjugate of $ababABaBABAB$) as a product of three commutators. Since extremal surfaces are π_1 -injective, the group generated by $abaB, \dots, aaBAAb$ is a free subgroup of F_2 of rank 6, equal to the image of π_1 of a genus 3 surface with one puncture under an injective homomorphism. This fact may be verified independently e.g. by using Stallings folding, and gives an independent check of the validity of the calculation.

As was discussed in the introduction, this example answers Bavard's question in the negative. Other simple examples are

$$\text{scl}(aababaBAaBAB) = 2/3$$

and

$$\text{scl}(a + BB + AAbab) = 3/4$$

4.1. Addendum. Several developments building on this material have taken place between the time this paper was submitted and was accepted for publication. It would be inappropriate to go into too much detail, but for the convenience of the reader, we summarize some of the most interesting points.

Firstly, the algorithm described above has been improved to run in *polynomial time*. An account of this improvement is described in [9], § 4.1.7–8. The program `scallop` (source code available at [8]) implements this algorithm, and can be used to compute scl in F_2 on words of length ~ 60 . Experiments using `scallop` reveal additional structure in the scl spectrum of a free group, partially explained in a forthcoming paper [7]. In particular, it is possible to prove rigorously that in any nonabelian free group, the image of scl contains nontrivial accumulation points, and takes on values with any denominator.

Secondly, the polyhedral structure on the scl norm unit ball is studied in [6], and it is shown that every realization of a free group F as $\pi_1(S)$ where S is an oriented surface with boundary is associated to a codimension one face of the boundary of the scl norm unit ball in B_1^H .

Thirdly, the extremal surfaces guaranteed in bullet (2) of the Rationality Theorem are exploited to construct closed surface subgroups in certain graphs of free groups amalgamated along cyclic subgroups. This is studied in [5], and further generalized in [13].

5. ACKNOWLEDGMENT

While writing this paper I was partially funded by NSF grant DMS 0707130. I would like to thank Roger Alperin, Nathan Dunfield, Dieter Kotschick, Lars Louder, Jason Manning, Bill Thurston, Dongping Zhuang, the anonymous referees, and the members of the dōsemi at the Tokyo Institute of Technology, especially Shigenori Matsumoto. I presented an incorrect argument, purporting to prove the main result of this paper, at the dōsemi in April 2007. Matsumoto asked a simple-sounding question about branch points, which turned out to be a crucial detail that I had overlooked. Understanding this detail was the key to obtaining the results in this paper. I would also like to explicitly thank Jason Manning for a number of conversations that substantially increased my confidence in the experimental results discussed in § 4.

REFERENCES

- [1] C. Bavard, *Longueur stable des commutateurs*, Enseign. Math. (2), **37**, 1-2, (1991), 109–150
- [2] A. Bouarich, *Suites exactes en cohomologie bornée réelle des groupes discrets*, C. R. Acad. Sci. Paris Sér. I Math. **320** (1995), no. 11, 1355–1359
- [3] R. Brooks, *Some remarks on bounded cohomology*, Riemann surfaces and related topics: Proceedings of the 1978 Stony Brook Conference (SUNY Stony Brook NY 1978) Ann. Math. Stud. **97**, Princeton Univ. Press 1981, 53–63
- [4] D. Calegari, *Foliations and the geometry of 3-manifolds*, Oxford Mathematical Monographs, Oxford Univ. Press, Oxford, 2007
- [5] D. Calegari, *Surface subgroups from homology*, Geom. Top. **12** (2008), 1995–2007
- [6] D. Calegari, *Faces of the scl norm ball*, Geom. Top. **13** (2009), 1313–1336
- [7] D. Calegari, *Scl, sails and surgery*, preprint, in preparation
- [8] D. Calegari, *scallop*, computer program, available from <http://www.its.caltech.edu/~dannyc>
- [9] D. Calegari, *scl*, monograph, to appear in Memoirs MSJ; available from <http://www.its.caltech.edu/~dannyc>
- [10] D. Calegari and K. Fujiwara, *Stable commutator length in word hyperbolic groups*, Groups, Geom. Dyn. to appear
- [11] G. Dantzig, *Linear Programming and Extensions*, Princeton Univ. Press, Princeton, 1963
- [12] D. Gabai, *Foliations and the topology of 3-manifolds*, J. Diff. Geom. **18** (1983), no. 3, 445–503
- [13] C. Gordon and H. Wilton, *On surface subgroups of doubles of free groups*, preprint, arXiv:0902.3693
- [14] M. Gromov, *Volume and bounded cohomology*, Inst. Hautes Études Sci. Publ. Math. No. 56 (1982), 5–99
- [15] M. Gromov, *Asymptotic invariants of infinite groups*, London Math. Soc. Lecture Note Ser. **182** Cambridge Univ. Press, Cambridge, 1993
- [16] J. Hempel, *3-Manifolds*, Ann. Math. Stud. **86**, Princeton Univ. Press, Princeton, 1976
- [17] M. Kiyomi, *exlp*, computer program, available from <http://members.jcom.home.ne.jp/masashi777/exlp.html>
- [18] A. Makhorin, *glpsol*, computer program, available from <http://www.gnu.org>

- [19] L. Mosher and U. Oertel, *Two-dimensional measured laminations of positive Euler characteristic*, Quart. J. Math. **52** (2001), 195–216
- [20] R. Penner with J. Harer, *Combinatorics of train tracks*, Ann. Math. Stud. **125**, Princeton Univ. Press, Princeton, 1992
- [21] U. Oertel, *Homology branched surfaces: Thurston's norm on $H_2(M^3)$* , Low-dimensional topology and Kleinian groups (Coventry/Durham, 1984), 253–272 London Math. Soc. Lecture Notes Ser. **112** Cambridge Univ. Press, Cambridge, 1986
- [22] R. Rannard, *Computing immersed normal surfaces in the figure-eight knot complement*, Experiment. Math. **8** (1999), no. 1, 73–84
- [23] P. Scott, *Subgroups of surface groups are almost geometric*, J. London Math. Soc. (2) **17** (1978), no. 3, 555–565
- [24] J.-P. Serre, *Trees*, Springer Monographs in Mathematics (corrected 2nd printing, trans. J. Stillwell), Springer-Verlag, Berlin, 2003
- [25] W. Thurston, *A norm for the homology of 3-manifolds*, Mem. Amer. Math. Soc. **59** (1986), no. 339, i-vi and 99–130
- [26] D. Zhuang, *Irrational stable commutator length in finitely presented groups*, Jour. Mod. Dyn. **2** (2008) no. 3, 499–507

DEPARTMENT OF MATHEMATICS, CALTECH, PASADENA CA, 91125

E-mail address: `dannyc@its.caltech.edu`